

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

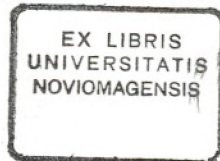
<http://hdl.handle.net/2066/74590>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

1 (1979)

Inhoudsoverzicht van Tijdschrift voor Taalbeheersing, 1979, jrg. 1.

Braet, A.C.	09
<i>Aspecten van de klassieke toep bij Aristoteles en Cicero (I)</i>	365-383
Dam, G. van	00
<i>De invloed van theoretische ontwikkelingen in de psychologie op de methoden van het onderzoek naar het onthouden van teksten</i>	61-74
Drop, W.	10
<i>Het argumentum ad hominem als tegenargument</i>	113-129
Dudink, A. en H. Meintjes	
<i>Kinderen aantekeningen laten maken</i>	261-272
Eemeren, F.H. van	
<i>Object en doelstelling op herhaling</i>	314-327
Geel, R.J.	
<i>De strijd om het opstel</i>	25-42
Haft-van Rees, M.A.	
<i>Het staat in de tekst; regels en conventies bij het interpreteren van teksten</i>	247-260
Hamaker, C.	
<i>Methodische problemen bij experimenteel tekstenmerken-onderzoek</i>	130-145
Jansen, C.J.M. en E.T. Woudstra	
<i>Theorie en praktijk van het Nederlandse leesbaarheidsonderzoek. Een analyse van twee formules</i>	43-60
Koning, W.K.B.	
<i>Geen taalbeheersing zonder sociolinguïstiek</i>	345-364
Looijmans, P.J.M. en J. Palm	
<i>Tekstbegrip getoetst?</i>	328-344
Palm, J.	
<i>Methodologische kanttekeningen bij de constructie en validering van leesbaarheidsformules</i>	180-192
Peeck, J.	
<i>Voorkennis en tekstbestudering</i>	301-313



Tz C
18677
T

21. In: Arthur Power, *Gesprekken met James Joyce*, Amsterdam 1976, blz. 119-120.
22. Handleiding, blz. 13.
23. op. cit. blz. 12.
24. op. cit. blz. 12.
25. op. cit. blz. 12.
26. op. cit. blz. 12.
27. Vgl. Derek Phillips, 'Enige beschouwingen over het wetenschappelijk onderwijs in Nederland'. *De Gids*, 140, nr. 9/10, 1977, blz. 621-623.
28. Advies, par. 2.3.1. Inhoudelijke relevantie.
29. K.L. Poll, 'Een Gericht Schrijven', NRC/Handelsblad, 28 oktober 1977.
30. Ik spreek mij er hier niet over uit of klassikaal opzeggen zinvol is. Het gaat er alleen om dat iets op school beoefend kan worden dat we niet in dezelfde vorm terugvinden in het 'normale leven'.
31. Advies, par. 2.3.2. Transparantie.
32. Vgl. mijn *Hoe zet ik mijn gedachten op papier*, Muiderberg 1977, tweede druk 1978, in het bijzonder het eerste hoofdstuk.
33. Advies, par. 2.3.4. Centraliteit van het examen.
34. Advies, par. 2.3.3. Betrouwbaarheid van de beoordeling.
35. NRC/Handelsblad, 11 november 1977.
36. Zie noot 35.
37. K.L. Poll, 'Drie grote vergissingen', NRC/Handelsblad 9 december 1977.
38. Leo Prick, 'Het eindexamen Nederlands: de stelopdracht'. *Levende Talen*, nr. 326, oktober 1977, blz. 385 e.v.

Bibliografie

Advies over het eindexamen Nederlands in het v.w.o. het h.a.v.o. en het m.a.v.o. Commissie modernisering moedertaalonderwijs, 's-Hertogenbosch april 1977.

C.M. Bolle, 'Weg met opstel op eindexamen'. NRC/Handelsblad, 30 september 1977.

W. Drop en J.H.L. de Vries, *Taalbeheersing*. Groningen 1974.

Rudolf Geel, *Hoe zet ik mijn gedachten op papier*. Muiderberg 1977.

Handleiding voor gericht schrijven, ontwikkeld uit ervaringen in scholen voor v.w.o. commissie modernisering moedertaalonderwijs. 's-Hertogenbosch 1976.

B. Meuffels, 'Effect-onderzoek taalvaardigheid', *Spektator* 8, nr. 1/2 1978-1979 blz. 47 e.v.

Derek Phillips 'Enige beschouwingen over het wetenschappelijk onderwijs in Nederland'. *De gids*, 140, nr. 9/10, 1977. blz. 615 e.v.

K.L. Poll, 'Een Gericht Schrijven'. NRC/Handelsblad, 28 oktober 1977.

K.L. Poll, 'Drie grote vergissingen'. NRC/Handelsblad 9 december 1977.

Arthur Power, *Gesprekken met James Joyce*. Amsterdam 1976.

Leo Prick, 'Het eindexamen Nederlands: de stelopdracht'. *Levende Talen*, nr. 326, oktober 1977, blz. 385 e.v.

Theorie en praktijk van het Nederlandse Leesbaarheidsonderzoek

Een analyse van twee formules*

C.J.M. JANSEN EN E.T. WOULDSTRA

Samenvatting

In dit artikel geven we eerst een kort overzicht van de stand van zaken in het Nederlandse leesbaarheidsonderzoek. Naar aanleiding daarvan bespreken we twee onlangs voor het Nederlands ontwikkelde leesbaarheidsformules, voorgesteld door Van Hauwermeiren (1975a). Daarbij gaan we, aan de hand van hoofdstuk 2 uit Van Hauwermeiren (1975b), in op de procedure die gevolgd is bij de constructie van die formules. Daarna presenteren we de resultaten van een kruisvalidatie-onderzoek. Onze conclusie luidt dat beide formules afgewezen dienen te worden. Tot slot worden nog enkele suggesties voor verder onderzoek gedaan.

1. Inleiding

In een artikel in *Spektator* getiteld 'Leesbaarheidsformules voor informatieve Nederlandse teksten' (Van Hauwermeiren 1975a) worden vier leesbaarheidsformules voorgesteld waarmee het mogelijk zou zijn de relatieve leesbaarheid¹ van informatieve Nederlandse teksten te bepalen. Over twee van deze formules zegt Van Hauwermeiren (1977) dat ze niet aan de statistische eisen voldoen die aan formules van dit type gesteld moeten worden²; op de andere twee formules, $C\% = 62.862 - 0.418X_7 - 0.323X_9$ en $C\% = 40.082 + 0.422X_2 - 0.391X_7 - 0.344X_9$ (verder aangeduid als resp. F_1 en F_2)³, zullen we in dit artikel ingaan. In paragraaf 2 zullen we deze twee formules bespreken aan de hand van een aantal methodologische beschouwingen uit *Het leesbaarheidsonderzoek* (Van Hauwermeiren 1975b). In paragraaf 3 zullen we opzet en resultaten van ons kruisvalidatie-onderzoek bespreken. In paragraaf 4 zal de conclusie worden gegeven en in paragraaf 5 zullen we enkele suggesties voor verder onderzoek doen.

In deze inleiding geven we daaraan voorafgaand achtereenvolgens een omschrijving van het begrip leesbaarheidsformule en een kort overzicht van de stand van zaken in het Nederlandse leesbaarheidsonderzoek⁴.

1.1 Wat is een leesbaarheidsformule?

Leesbaarheidsformules zou men kunnen omschrijven als instrumenten waarmee men de (relatieve of absolute) leesbaarheid van teksten kan voorspellen op grond van de waarden die twee of meer linguïstische variabelen in (een fragment uit) die teksten bezitten.

Leesbaarheidformules hebben meestal de vorm van multiple lineaire regressievergelijkingen: wiskundige formules met als vorm: $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$, waarin Y staat voor de voorspelde leesbaarheid, a en b_i tot en met b_n constanten zijn, en X_1 tot en met X_n de waarden van linguïstische variabelen (als bijv. gemiddelde woord- of zinslengte) in de tekst vertegenwoordigen.

Hoe een leesbaarheidsformule tot stand komt is o.a. beschreven in Van Hauwermeiren (1975b) en (1977). Beknopt samengevat gaat het om de volgende procedure:

- Bepaal de leesbaarheid van een aantal tekstfragmenten door een aantal proefpersonen daarover een leestest af te nemen.
 - Bepaal de waarde van een aantal linguïstische variabelen in die fragmenten.
 - Bereken op grond van de gegevens uit de eerste twee fasen de coëfficiënten a en b_1 tot en met b_n en stel daarmee de formule samen.
- Bezien we nu de wijze waarop tot nu toe Nederlandse leesbaarheidsformules tot stand zijn gekomen, dan zijn daartegen een aantal bezwaren aan te voeren. In deze paragraaf zullen we twee overzichtsartikelen bespreken, waaruit duidelijk wordt dat de meeste Nederlandse formules erg zwak onderbouwd zijn. In paragraaf 2 en 3 willen we laten zien dat dat ook geldt voor F_1 en F_2 .

1.2 De stand van zaken in het Nederlandse leesbaarheidsonderzoek

Nog tamelijk recent zijn er twee artikelen verschenen waarin een overzicht werd gegeven van het Nederlandse leesbaarheidsonderzoek: Brouwer (1976) en Van Hauwermeiren (1977).

a Brouwer maakt in zijn kritische bespreking een onderscheid tussen *vernederlandste Flesch-formules*, de *Cloze procedure van Taylor* en *leesbaarheidsformules leesindex A, B en AB*. In de eerste categorie bespreekt hij de eerste toepassingen van de originele Flesch-formules in het Nederlands door B. de Leeuw-Polak (in 1957) en door T. Stoelen (in 1958). Omdat al snel bleek dat de Flesch-formules niet zonder meer voor het Nederlands konden worden gebruikt⁵, werden 'gecorrigeerde' Flesch-formules ontwikkeld: de formules van Douma en van Glimmerveen/Van der Werff. Van beide formules maakt Brouwer duidelijk, dat de criteria waarop de aanpassingen uit het Engels zijn gebaseerd, uitermate zwak zijn. Zo geeft hij het argument waarop Douma afrondingen van 10% in zijn formule rechtvaardigt: een mededeling van de hoofdredacteur van *Het Beste*, dat deze op grond van ervaring rekening moest houden met 10% uitloop bij het opnemen van vertaalde artikelen uit *Readers Digest*.

Naar aanleiding van de cloze procedure⁶ merkt Brouwer, overigens ten onrechte⁷, op dat die nog niet als basis heeft gefungeerd voor leesbaarheidsformules voor het Nederlands.

Met betrekking tot de leesbaarheidsformules leesindex A, B en AB tenslotte, die Brouwer zelf heeft ontwikkeld, merkt hij op dat er tot nu toe alleen kritiek is geleverd op onderdelen van het onderzoek naar de leesindex A-formule, waarna hij die kritiek tracht te weerleggen.

b Ook Van Hauwermeiren (1977) bespreekt eerst de vernederlandste Flesch-formules en staat er even afwijzend tegenover als Brouwer. Daarnaast echter wijst hij ook de formules van Brouwer van de hand, "[...] omdat leesbaarheids-predictie, die niet op leesbaarheidsmeting stoelt, geen enkele statistische waarde heeft" (p. 165)⁸.

Verder behandelt Van Hauwermeiren zg. oorspronkelijke Nederlandse formules

en wel: a. de vier formules uit Van Hauwermeiren (1975a), waarvan hij zoals eerder opgemerkt, er twee afwijst; b. de formules van Luyten-Sijben, waarvan hij opmerkt: "Statistisch gezien zijn deze beide formules ondeugdelijk" (p. 173); en c. de formules van Zondervan/Van Steen/Gunneweg, die de leesbaarheid van basisschoolteksten voorspellen voor leerlingen van één klas van de basisschool. Ook op deze laatste formules heeft Van Hauwermeiren kritiek zij het onzes inziens van minder fundamenteel karakter. Zo zegt hij o.a. dat de basis waarop de proefpersonen zijn gekozen, nl. de genoten opleiding, een verkeerde is, als men er tenminste zeker van wil zijn dat de leesvaardigheid van de proefpersonen even groot is (p. 176).

Zowel de kritische opmerkingen van Brouwer (1976) als die van Van Hauwermeiren (1977) lijken ons terecht, tenminste voor zover het niet de door de auteurs zelf vervaardigde formules betreft. Dat de formules van Brouwer minder waardevol zijn dan uit diens overzichtsartikel blijkt, lijkt ons door Van Hauwermeiren afdoende te zijn aangetoond; dat aan F_1 en F_2 , de enige twee formules die van Hauwermeiren in zijn artikel niet van de hand wijst, ook het een en ander schort, willen we in de volgende twee paragrafen laten zien.

2. Enkele methodologische beschouwingen bij F_1 en F_2

In deze paragraaf zullen we eerst een korte samenvatting geven van Van Hauwermeiren (1975a), waarin hij F_1 en F_2 voorstelt. Daarna zullen we de constructie van F_1 en F_2 bespreken aan de hand van het tweede hoofdstuk uit Van Hauwermeiren (1975b) getiteld: *Methodologische beschouwingen bij de meting en de predictie van de leesbaarheid*.

2.1 Een korte samenvatting van Van Hauwermeiren (1975a)

De vier leesbaarheidsformules die Van Hauwermeiren hier voorstelt, en waarmee het mogelijk zou zijn te bepalen "[...] hoe begrijpelijk informatieve teksten zijn voor lezers met middelbare-schoolopleiding" (p. 500) luiden als volgt:

$$C\% = 62.862 - 0.418X_1 - 0.323X_9 \quad (F_1)$$

$$C\% = 40.082 + 0.422X_2 - 0.391X_7 - 0.344X_9 \quad (F_2)$$

$$C\% = 136.067 - 0.462X_1 + 1.535X_3 - 0.479X_7 - 0.364X_9$$

$$C\% = 258.767 - 1.125X_1 + 3.915X_3 - 0.778X_4 - 0.605X_9 - 1.331X_{13} + 1.717X_{17}$$

waarin $C\%$ = het voorspelde clozepercentage, X_1 = lettergrepen per honderd woorden, X_2 = eenlettergrepige woorden per honderd woorden, X_3 = woorden met meer dan drie lettergrepen per honderd woorden, X_4 = woorden per zin, X_7 = lettergrepen per T-eenheid, X_9 = woorden in de tangconstructies per honderd woorden, X_{13} = voorzetsels per honderd woorden en X_{17} = infinitiefconstructies per honderd woorden.

Voor de constructie van de formules heeft Van Hauwermeiren gebruik gemaakt van de uitslagen van cloze procedure tests, uitgevoerd door Van Remoortele (1974) in het kader van een onderzoek naar de factoren die het nauwste verband vertonen met de leesbaarheid van uit het Engels vertaalde populair-

wetenschappelijke teksten. Het ging bij die cloze procedure tests om 16 tekstfragmenten van ongeveer 200 woorden lang en om 120 proefpersonen, allen leerlingen van het voorlaatste en laatste jaar oude en moderne humaniora in België (vergelijkbaar met het voorlaatste en laatste jaar vwo en havo in Nederland). Van Hauwermeiren bepaalde de waarde van 20 linguïstische variabelen in de betreffende fragmenten en construeerde met behulp van correlatie- en regressierekening de vier formules. Ten slotte werd nog een factoranalyse uitgevoerd, die als de twee belangrijkste factoren 'syntactische lengte' en 'morfologische lengte' opleverde.

2.2 Een analyse van F_1 en F_2

Van Hauwermeiren (1975b) biedt een uitstekend overzicht van het leesbaarheidsonderzoek tot dan toe, met name van wat het Engelse taalgebied betreft. In het tweede hoofdstuk van die publikatie worden een aantal aspecten genoemd waarmee men bij de constructie en evaluatie van leesbaarheidsformules rekening dient te houden. We willen hier bij de bespreking van F_1 en F_2 laten zien aan welke van de daar genoemde aspecten niet voldoende aandacht is besteed. We zullen daarbij achtereenvolgens behandelen: het gehanteerde criterium, de keuze van de proefpersonen, de correlatie tussen de variabelen en de operationeel gedefinieerde leesbaarheid, het opstellen van de handleiding, de interpretatie van de uitslagen, de betrouwbaarheid in het gebruik van de formules, en tenslotte de validiteit van de formules⁹.

2.2.1 Het gehanteerde criterium

Zondervan en Van Caldenborgh (1971) concluderen, na een bespreking van de complexe relatie tussen leesbaarheid als object-construct (het begrip zoals bedoeld) en de resultaten van leesbaarheidsformules (voorspellingen over de waarde die het gehanteerde criterium in te onderzoeken teksten zal aannemen), dat de voor de hand liggende uitspraak dat leesbaarheidsformules de leesbaarheid van een tekst meten, als een zinledige uitspraak moet worden beschouwd (p. 350). Het belangrijkste argument voor die conclusie is dat geen van de mogelijke criteria die men kan hanteren, het object-construct geheel kan dekken, omdat het begrip "[...] onvermijdelijkerwijze een *surplus-betekenis* [heeft] ten opzichte van ieder empirisch criterium" (De Groot 1966: 273).

In het geval van leesbaarheidsmeting en -voorspelling lijkt ons het door De Groot gesignaleerde probleem extra groot. Van een enigszins ontwikkelde theorie over tekstbegrip is immers (nog) geen sprake: een algemeen aanvaarde definitie van tekstbegrip ontbreekt nog. Het gevolg is dat alleen de voorwetenschappelijke intuïties van de onderzoekers die op dit terrein werkzaam zijn, bepalend zijn voor de acceptabiliteit van empirische leesbaarheidscriteria: een niet erg bevredigende situatie (zie ook Jansen en Woudstra 1978).

Om weer terug te keren naar F_1 en F_2 : ook als deze formules goede voorspellers zouden blijken voor het gemiddeld percentage goed gescoorde cloze-items (het hier gehanteerde criterium), dan nog blijft de vraag in hoeverre de formules iets zeggen over de leesbaarheid van teksten.

In Van Hauwermeiren (1975a) komt deze vraag niet aan de orde, maar wordt zonder meer gesteld dat de formules ons inlichten over de (relatieve)

leesbaarheid van informatieve teksten. In Van Hauwermeiren (1975b) komt het probleem wel ter sprake. In par. 2.3.2 bespreekt hij de tot nu toe meest gebruikte criteria voor leesbaarheid: het percentage juist beantwoorde vragen die naar aanleiding van teksten worden gesteld, en de gemiddelde score behaald bij een cloze procedure test. Tegen beide criteria noemt hij een aantal bezwaren en hij concludeert (p. 88): "De twee manieren waarop tot nu toe gemeten werd wat de proefpersonen van de testpassages begrijpen, vertonen dus beide tekortkomingen". Alleen al op grond van deze conclusie lijkt het wat voorbarig om cloze procedure score-verschillen zonder meer te interpreteren als verschillen in de (relatieve) leesbaarheid.

Maar ook als we de cloze procedure scores als een bruikbaar criterium beschouwen, blijven er bij ons problemen bestaan met betrekking tot: (a) de teksten die Van Hauwermeiren als criteriumteksten heeft gebruikt, en (b) de manier waarop de cloze procedure test volgens Van Remoortele (1974) is afgenomen. Op die bezwaren zullen we hieronder ingaan.

a De criteriumteksten

Van Hauwermeiren (1975b:73) noemt twee eisen waaraan criteriumteksten moeten voldoen. De eerste eis is "[...] verschil in leesbaarheid [...], omdat het precies dit verschil is dat de variabelen in de taal van de passages geacht worden te weerspiegelen. Dat leesbaarheidsverschil mag niet te klein zijn, want een formule zal geen betrouwbare informatie kunnen verstrekken over de leesbaarheid van teksten die gemakkelijker zijn dan de gemakkelijkste en moeilijker dan de moeilijkste tekstpassages die bij het ontwerpen van die formule zijn gebruikt". De tweede eis is, dat de criteriumteksten representatief moeten zijn voor het teksttype waarover de formule geacht wordt voorspellingen te doen. Naar aanleiding van de eerste eis nu, valt op te merken "[...] dat de gemiddelde cloze test-uitslagen slechts een geringe spreiding vertonen" (Van Hauwermeiren 1975a:501), zodat onzes inziens betwijfeld kan worden of het leesbaarheidsverschil in de criteriumteksten nog groot genoeg is.¹⁰

Naar aanleiding van de tweede eis – de representativiteit – rijzen bij ons ook een aantal vragen. Niet geheel duidelijk is voor welk type teksten de formules geacht worden te gelden. Op p. 499 en p. 500 spreekt Van Hauwermeiren over *informatieve (Nederlandse) teksten*, op p. 501/502 over *populariserend informatief proza* – naar onze mening slechts een deelverzameling van de eerste categorie. Daarmee wordt de vraag naar de representativiteit natuurlijk moeilijk te beantwoorden. Ook Van Remoortele (1974), die de 16 fragmenten selecteerde, maakt overigens melding van het representativiteitsprobleem. Zo zegt hij (p. 31): "[...] the writer is fully aware of the fact that many objections can be made to this choice [...]". Die bezwaren zouden dan bijv. betrekking kunnen hebben op zijn besluit (populair-) wetenschappelijke teksten over filosofie, literatuur, filologie en linguïstiek niet in de steekproef op te nemen.

Voorts lijkt ons met betrekking tot de representativiteit problematisch dat alle criteriumteksten vertalingen zijn uit het Engels; Van Hauwermeiren (1977: 163) noemt als bezwaar tegen de formules van Flesch-Douma, dat de gebruikte Nederlandse teksten alle vertalingen zijn en in een vroegere publikatie (Van Hauwermeiren 1963: 54) zegt hij: "Vertalingen nu, bieden het nadeel dat stijlelementen als zinslengte enz. in grote mate gelijkemaakt zijn aan de

vertaalde tekst". Het lijkt ons dat ook op grond hiervan de keuze van de 16 teksten niet erg gelukkig genoemd kan worden, gezien ook de belangrijke rol van de variabele X_9 (woorden in de tangconstructies per honderd woorden) in de formules. In Van Hauwermeiren (1975a:507) immers zegt hij dat de tangvorming eigen is aan het Nederlands. Juist wanneer men gebruik maakt van vertalingen lijkt ons het gevaar reëel dat de waarden die men voor een variabele als X_9 vindt, een vertekend beeld geven¹¹.

b De manier waarop de cloze procedure is afgenomen

Van Hauwermeiren (1975b: 81-88) bespreekt literatuur waarin het construeren, toedienen, scoren en interpreteren van de cloze procedure aan de orde komt. We willen hier nader ingaan op het verloop van twee fasen uit die procedure in het onderzoek beschreven in Van Hauwermeiren (1975a), nl. het construeren van de test en het toedienen van de test.

Het construeren van de test. Vooral op grond van Bormuth (1964) suggereert Van Hauwermeiren (1975b: 82) om tekstpassages van ten minste 250 woorden te gebruiken, elk vijfde woord weg te laten, en over elke tekstpassage vijf verschillende tests af te nemen: een test waarin het vijfde, het tiende, het vijftiende, . . . enz. woord wordt weggelaten, een test waarin het zesde, het elfde, het zestiende, . . . enz. woord wordt weggelaten, enz. Die laatste maatregel is bedoeld om te voorkomen dat de ene tekstpassage bevoordeeld zou worden boven de andere, doordat de verhouding in moeilijkheid tussen weggelaten en niet-weggelaten woorden voor de ene tekstpassage veel gunstiger zou zijn dan voor de andere.

Zien we nu hoe Van Remoortele (1974) zijn cloze procedure test heeft geconstrueerd, dan valt het volgende op:

- Alle fragmenten bestaan uit minder dan 250 woorden; het kortste fragment bevat zelfs maar 142 woorden. Waarom is niet duidelijk.
- Over elke tekstpassage zijn inderdaad vijf tests afgenomen, maar uit Van Remoortele (p. 37) blijkt dat bij het vervaardigen van de testversies eigennamen in bepaalde gevallen buiten beschouwing zijn gelaten. Steeds namelijk wanneer een eigenaam voor deletie in aanmerking kwam, omdat die eigenaam het vijfde woord was na de voorafgaande deletie, werd niet die eigenaam, maar blijkens de voorbeelden in de appendix van Van Remoortele (1974), het woord onmiddellijk daaropvolgend gedeleerd. Vervolgens werd steeds weer vijf woorden doorgeteld om de volgende deletie aan te brengen. Deze procedure nu had tamelijk vervelende gevolgen, vooral wanneer een tekst in het begin veel eigennamen bevatte. De procedure leidde er dan namelijk toe dat vanaf een bepaald punt in de tekst twee of meer van de vijf tekstversies niet meer van elkaar verschilden. Daarmee werd uiteraard bevoordeling van de ene tekstpassage boven de andere niet langer uitgesloten. Een goede oplossing lijkt ons de eigennamen bij het vervaardigen van de testversies *helemaal* niet in de telling te betrekken, bijvoorbeeld door ze tussen haakjes te zetten voordat met het vaststellen van de deleties wordt begonnen.

Het toedienen van de test. Van Remoortele heeft zijn 120 proefpersonen verdeeld in 20 groepen van 6, waarbij elke groep 4 verschillende testversies kreeg toebedeeld, die zo gekozen waren dat het verschillende versies van verschillende teksten waren (Van Hauwermeiren 1975a: 501).

Niet geheel duidelijk daarbij is of de 6 proefpersonen die één groep vormden hun 4 testversies steeds in dezelfde volgorde kregen aangeboden, maar op grond van het distributieschema dat Van Remoortele (p. 36) geeft, ligt het voor de hand om aan te nemen dat dat inderdaad het geval was. Dat zou dan onzes inziens wel een bedreiging voor de kwaliteit van het onderzoek betekenen, omdat de proefpersonen dan steeds versies van dezelfde teksten als laatste aangeboden kregen. Factoren als vermoeidheid, verzadiging en tijdgebrek zouden dan immers storende variabelen kunnen zijn.¹²

2.2.2 De keuze van de proefpersonen

Een eerste eis die men aan de keuze van proefpersonen bij het vervaardigen van een leesbaarheidsformule moet stellen, lijkt ons dat de proefpersonen representatief zijn voor de populatie waarover de formule geacht wordt voorspellingen te doen. Een tweede eis die men zou moeten stellen is dat de leesvaardigheid van de proefpersonen bekend is, voordat men de testpassages over de proefpersonen verdeelt.

Naar aanleiding van de eerstgenoemde eis valt op te merken dat onduidelijk is, waarom de gekozen groep proefpersonen, 120 leerlingen van een middelbare school, representatief zou zijn voor *lezers met een middelbare-schoolopleiding* (Van Hauwermeiren 1975a: 500). Een reden om daaraan te twifelen geeft Van Hauwermeiren (1975b: 90), als hij de onderzoeksresultaten noemt van Norris en van Lorge en Blau, die melding maken van 'a quite general improvement among adults in the ability to understand and use the English language'. We vragen ons af of ook niet voor het Nederlands zou gelden dat leerlingen van een middelbare school een geringere leesvaardigheid bezitten dan mensen die eenzelfde opleiding al enige tijd achter de rug hebben. Als dat zo zou zijn, is de generalisering van Van Hauwermeiren over de groep lezers met een middelbare-schoolopleiding niet gerechtvaardigd.

Naar aanleiding van de tweede eis die we hierboven hebben genoemd, valt op te merken dat uit Van Hauwermeiren (1975a), noch uit Van Remoortele (1974) geconcludeerd kan worden dat over de individuele leesvaardigheid van de proefpersonen iets bekend was. Dat betekent dat een storende invloed van deze variabele in het onderzoek heel wel mogelijk is geweest. Het is niet uitgesloten dat de ene tekstpassage is bevoordeeld boven de andere doordat de groep proefpersonen die de cloze procedure test over de ene tekstpassage uitvoerde gemiddeld leesvaardiger was dan de groep proefpersonen die zich met de andere tekstpassage bezig hield. Op een dergelijk gevaar wordt overigens ook gewezen in Van Hauwermeiren (1977: 176) waar hij opmerkt, dat men er rekening mee moet houden "[...] dat proefpersonen uit eenzelfde leer- of studiejaar daarom nog niet dezelfde leesvaardigheid hebben".

2.2.3 De correlatie tussen de variabelen en de operationeel gedefinieerde leesbaarheid

Van Hauwermeiren (1976a) geeft niet expliciet aan waarom hij van alle mogelijke multiple lineaire regressievergelijkingen die hij op grond van zijn (correlatie-)gegevens had kunnen afleiden, slechts de vier voorgestelde heeft gekozen. Het lijkt er echter op dat hij bij zijn beslissing de volgende criteria heeft gehanteerd:

- De linguïstische variabelen moeten in lineaire relatie staan tot het leesbaarheids criterium.
- De linguïstische variabelen moeten onderling lineair onafhankelijk zijn. (Dit criterium, evenals het eerstgenoemde, vloeit voort uit de aanname van een lineair model voor de beschrijving van het verband tussen leesbaarheid en linguïstische variabelen¹³, zie bijv. Blalock (1972: 456)).
- De (gecorrigeerde) multiple R^2 (een maat voor het verband tussen de scores behaald op de criteriumteksten en de door de formule voor die teksten voorspelde score) moet bevredigend groot zijn¹⁴.
- De formule moet zo hanteerbaar mogelijk zijn; een formule bijv. met 15 moeilijk telbare linguïstische variabelen is ongewenst.

Het is ons niet duidelijk waarom een formule bijv. met als variabelen X_1 (lettergrepen per honderd woorden), X_7 (lettergrepen per T-eenheid) en X_9 (woorden in de tangconstructies per honderd woorden) door Van Hauwermeiren niet is voorgesteld. Blijkens de correlatiematrix (Van Hauwermeiren 1957a: 508) zou in zo'n formule aan de twee eerstgenoemde criteria worden voldaan; de (gecorrigeerde) multiple R^2 bedraagt 0.746 – nauwelijks minder dan die van F_2 (0.757) en meer dan die van F_1 (0.651). Wat het laatstgenoemde criterium betreft tenslotte, de hanteerbaarheid van een dergelijke formule zou in ieder geval groter geweest zijn dan die van F_2 : het tellen van het aantal eenlettergrepige woorden zou onnodig zijn geworden.

2.2.4 Het opstellen van de handleiding

Een van de eisen waaraan een handleiding bij een leesbaarheidsformule zou moeten voldoen volgens Van Hauwermeiren (1975b: 96) is dat er gebruiksaanwijzingen in te vinden zijn voor het berekenen van de waarden van de linguïstische variabelen in de te onderzoeken passages.

Een gebruiker nu van F_1 en F_2 heeft als handleiding voor het berekenen van de waarden van de linguïstische variabelen alleen Van Hauwermeiren (1975b: 502-506) tot zijn beschikking. Met name de richtlijnen die daar worden gegeven voor het berekenen van X_7 (lettergrepen per T-eenheid) en X_9 (woorden in de tangconstructies per honderd woorden) plaatsen ons in een groot aantal gevallen voor problemen die ernstiger zijn dan zou blijken uit een mededeling als: "Het percentage woorden in de tangconstructie is een voorbeeld van een variabele die van de formulegebruikers een minimum aan grammaticale kennis en routine in het ontleden veronderstelt en daardoor de formule moeilijk om toe te passen maakt" (Van Hauwermeiren 1977: 178)¹⁵.

a De T-eenheid

Een T-eenheid omschrijft Van Hauwermeiren (1975a: 503) als volgt: "Zowel de onafhankelijke zin als de verbinding van een hoofdzin en zijn bijzin(nen) noemen we een *T-eenheid* (Eng. *Terminable unit*)"; bovendien stelt hij: "Een zin die bestaat uit één enkele subject-predikaatsverbinding is een onafhankelijke zin; een zin die bestaat uit twee of meer subjects-predikaatsverbindingen die samen een hiërarchische structuur vormen is een samengestelde zin met een hoofdzin en één of meer bijzinnen". Deze omschrijving is voor ons niet voldoende duidelijk om bijv. te beslissen of een geval van samentrekking als *Jan koopt een jas en Piet een hoed* als één T-eenheid of als twee T-eenheden moet worden opgevat. Over de

vraag of het hier gaat om één of twee subject-predikaatsverbindingen zullen niet alle grammatici het eens zijn. Voor een leek is dit in ieder geval een probleem. Mocht bedoeld zijn dat het in een dergelijk geval om twee T-eenheden gaat, dan is nog niet duidelijk hoeveel T-eenheden een zin als *Jan koopt een jas en een hoed* bevat. Ook hier immers lijkt er iets voor te zeggen de zin een dieptestructuur met nevenschikking op S-niveau toe te denken. De handleiding is op dit punt niet precies genoeg.

b De tangconstructies

Tangconstructies komen volgens Van Hauwermeiren (1975a: 503-505) voor in a) hoofdzinnen, b) bijzinnen, c) beknopte bijzinnen met een infinitief als kern, en d) als substantiefgroep van het type 'determinator + voorbepaling + substantief'. In de gevallen b) en c) levert het tellen van de woorden in de tang geen problemen op: de eerste en de tweede grijper (in de (gewone) bijzin resp. onderwerp en persoonsvorm, en in de beknopte bijzin met een infinitief als kern resp. voorzetsel en te + infinitief) zijn gemakkelijk te onderscheiden. In de gevallen a) en d) is het moeilijker de twee grippers te onderscheiden.

De hoofdzin In de hoofdzin is een van de twee grippers altijd de persoonsvorm. De andere – en daar beginnen de problemen – kan zijn volgens Van Hauwermeiren: het niet-werkwoordelijk deel van een werkwoordelijke uitdrukking, de infinitief die deel uitmaakt van het werkwoordelijk gezegde, het voltooid deelwoord dat deel uitmaakt van het werkwoordelijk gezegde, het naamwoordelijk deel van het gezegde, het specificerend complement, de predikatieve bepaling of het complement van *maken* en *vinden*, het direct object, het voorzetselvoorwerp, de aanvulling bij werkwoorden die een voortbeweging uitdrukken en ten slotte: de aanvulling bij werkwoorden die een verblijven uitdrukken. Het meest problematisch aan deze opsomming is dat een niet- of weinig taalkundig geschoolde formulegebruiker hier helemaal niet mee uit de voeten kan. De gebruikte begrippen zijn zo specialistisch dat zelfs moeilijk voorstelbaar is hoe een taalkundig wel geschoolde gebruiker, die bijv. minder goed thuis is in Klooster e.a. (1970) dan in Rijpsma en Schuringa (1969), de formules efficiënt zou kunnen toepassen.

Ook echter wanneer men wel thuis is in het gebruikte begrippenapparaat blijven er problemen. Om er een te noemen: Van Hauwermeiren (1975a: 503) geeft als voorbeeld van een tangconstructie waarin de tweede grijper zou moeten zijn de infinitief die deel uitmaakt van het werkwoordelijk gezegde, de volgende zin: ik zal nu toch wel niet langer snijbonen *moeten* eten. Blijkens de cursivering wordt hier *moeten* als de tweede grijper beschouwd. Onduidelijk is voor ons waarom niet *eten* als zodanig zou kunnen gelden, vooral gezien de opmerking van Van Hauwermeiren op p. 505/506: "Bij het tellen van de 'woorden in de tangconstructie' hebben we geen rekening gehouden met de 'tang in de tang'. We hebben alle woorden in de eerste tang slechts eenmaal geteld ook als ze in een tweede of derde tang voorkwamen." Hieruit zou kunnen worden opgemaakt dat steeds de grootst mogelijke tang moet worden gezocht.

De substantiefgroep Ook de substantiefgroep van het type 'determinator + voorbepaling + substantief' beschouwt Van Hauwermeiren als een tangconstructie, waarin de determinator en het substantief de grippers van de

tang zijn. Daarbij moeten de woorden in de tang wel "[...] duidelijk een beknopte bijzin vormen, die tussen de determinator en het substantief is geplaatst [...]" (p. 505).

Deze omschrijving stelt ons voor twee problemen. Het eerste is: geldt een substantiefgroep als *de te behandelen constructies* nu als tangconstructie, of is een voorbepaling als *te behandelen* niet even duidelijk een beknopte bijzin als bijv. *in dit artikel te behandelen* in de *in dit artikel te behandelen constructies*? Het tweede probleem betreft het begrip determinator: geldt bijv. *in dit artikel behandelde constructies* als een tangconstructie met als determinator het lidwoord 'nul' (zie bijv. Klooster e.a. 1970: 22), of is hier geen sprake van een lidwoord, noch van een determinator, zodat er ook geen sprake kan zijn van een tangconstructie? Voor beide problemen is in de richtlijnen noch in de voorbeelden van Van Hauwermeiren (1975a) een oplossing te vinden.

Concluderend: de handleiding bij F_1 en F_2 laat te veel problemen onopgelost om nog bevredigend genoemd te kunnen worden.

2.2.5 De interpretatie van de uitslagen

"De toepassing van een leesbaarheidsformule leidt tot een uitslag die eerst dan praktisch bruikbaar is als hij kan worden geïnterpreteerd, d.i. als kan worden bepaald met welk leesvaardigheidsniveau hij correspondeert," aldus Van Hauwermeiren (1975b: 96). Helaas ontbreekt in Van Hauwermeiren (1975a) een schaal waarmee een dergelijke interpretatie mogelijk zou zijn. Wel geeft Van Hauwermeiren (1975b: 97-111) een ingenieuze procedure om langs empirische weg zo'n schaal te construeren.

Uitgangspunt van deze procedure is, dat een tekst geschikt is voor een lezer als hij door het lezen ervan méér dan een bepaald percentage *informatiewinst* behaalt. De informatiewinst kan men vaststellen door de lezer vóór en na het lezen een (bijv. cloze-) test af te nemen; het verschil tussen de scores – dat gecorrigeerd wordt om regressie naar het gemiddelde (zie bijv. Mellenbergh 1976: 27-30) te vermijden – is de informatiewinst. Wanneer de tekst erg gemakkelijk is, zal de informatiewinst gering zijn, omdat de lezer al vóór het lezen van de tekst hoog zal scoren op een (cloze-)test daarover. Wanneer de tekst erg moeilijk is, zal de informatiewinst evenzeer zeer gering zijn, omdat zowel voor als na het lezen van de tekst de score op de (cloze-)test laag zal zijn. Alleen als de tekst noch te gemakkelijk, noch te moeilijk is, kan men een hoge informatiewinst verwachten. Heeft men nu een reeks teksten waarvan de formule-uitslag bekend is en legt men die voor aan proefpersonen waarvan de leesvaardigheid bekend is, dan kan men vaststellen bij welke formule-uitslag(en) de informatiewinst van proefpersonen met een bepaalde leesvaardigheid boven een door de onderzoeker te kiezen minimum uitgaat. Teksten met die formule-uitslag(en) zijn dus 'geschikt' voor lezers met de betreffende leesvaardigheid. Het lijkt zinnig om deze procedure in de praktijk op haar waarde te toetsen.

Het is jammer dat in Van Hauwermeiren (1975a) alleen met een voortoets is gewerkt en niet ook met een natoets. Dan immers zou de informatiewinst berekend kunnen worden en zou een eerste aanzet voor het opstellen van goede interpretatieschalen mogelijk zijn geweest. De praktische bruikbaarheid van F_1 en F_2 zou daarmee zijn toegenomen.

2.2.6 De betrouwbaarheid in het gebruik van de formules

De betrouwbaarheid in het gebruik van een leesbaarheidsformule wordt volgens Van Hauwermeiren (1975a: 111) bepaald door de betrouwbaarheid van 1) de steekproeftrekking (uit de door de gebruiker te onderzoeken teksten) en 2) de analyse (van diezelfde teksten).

De betrouwbaarheid nu van de analyse – de mate waarin twee of meer analyses van dezelfde steekproef overeenkomen, door dezelfde of verschillende formulegebruikers verricht – is o.a. afhankelijk van de duidelijkheid en de precisie van de handleiding bij de formule. Gezien onze bezwaren tegen de handleiding bij F_1 en F_2 lijkt het ons dat de betrouwbaarheid van de analyse, en daarmee de betrouwbaarheid in het gebruik van F_1 en F_2 , onmogelijk voldoende kan zijn; de resultaten genoemd in paragraaf 3 bevestigen die veronderstelling.

2.2.7 De validiteit van de formules

Van Hauwermeiren (1975b) behandelt bij zijn bespreking van de validiteit van een leesbaarheidsformule (p. 118-122) de contaminatie van het criterium en de controlevaliditeit, waarbij hij weer een onderscheid maakt in vergelijkende validiteit, kruisvalidatie, validatie tegen een extern of onafhankelijk criterium en experimentele validatie. Problemen met betrekking tot begripsvaliditeit in het leesbaarheidsonderzoek behandelt hij hier niet. Voor onze opmerkingen over die problematiek: zie paragraaf 2.2.1. In deze paragraaf zullen we alleen nader ingaan op de contaminatie van het criterium. Met betrekking tot de controlevaliditeit willen we volstaan met op te merken dat Van Hauwermeiren (1975a en 1977) geen onderzoek met F_1 en F_2 op dat terrein noemt. Van het kruisvalidatie-onderzoek dat wij met F_1 en F_2 hebben uitgevoerd, vindt u het verslag in paragraaf 3.

De contaminatie van het criterium

Van Hauwermeiren (1975b, p. 118) haalt Chall (1958) en Klare (1963) aan als hij zegt dat “[...] we de validiteit van een formule enigszins kunnen afleiden uit de correlatie tussen de testcores voor de passages die bij het opstellen van de formule als criterium gefungeerd hebben en de formulescores voor dezelfde passages”, de multiple R dus die in paragraaf 2.2.3 – toen gekwadrateerd en gecorrigeerd voor het aantal vrijheidsgraden – al aan de orde is geweest. De multiple R voor F_1 en F_2 bedraagt resp. 0.836 en 0.898, tamelijk hoge waarden als men weet dat de multiple R kan lopen van 0 (geen lineair verband) tot 1 (volmaakt lineair verband). Toch moet men aan deze gegevens niet al te veel waarde toekennen, gezien de opmerking van Klare, die volgens Van Hauwermeiren (1975b: 118) zegt “[...] dat de correlatie een aanduiding van weinig waarde is, omdat de variabelen meestal gekozen worden omwille van hun hoge correlatie met het criterium”. Met andere woorden: een hoge multiple R is een noodzakelijke voorwaarde voor de validiteit van een leesbaarheidsformule, maar zeker geen voldoende.

3. Een kruisvalidatie-onderzoek

“De *kruisvalidatie* bestaat erin dat men de test op kleinere schaal overdoet, maar wat testpassages en proefpersonen betreft, een andere steekproef trekt. Hoe hoger de correlatie is tussen de testcores (waargenomen scores) en de

formulescores voor de testpassages (voorspelde scores) des te minder acht men de correlatie tussen het criterium van de leesbaarheid en de variabelen in de formule door toevalligheden bepaald". Aldus omschrijft Van Hauwermeiren (1975b: 120) het begrip kruisvalidatie¹⁶.

In het validatie-onderzoek dat we met betrekking tot F_1 en F_2 hebben uitgevoerd, zijn we op twee punten van bovenstaande omschrijving afgeweken: – de test is niet op kleinere schaal afgenomen, maar er zijn evenveel testpassages en evenveel proefpersonen gebruikt en – wat de testpassages betreft is niet een andere, maar dezelfde steekproef getrokken. Beide maatregelen zouden echter eerder tot een hogere dan tot een lagere correlatie tussen voorspelde en waargenomen scores moeten leiden.

De 120 proefpersonen die aan dit validatie-onderzoek hebben meegewerkt, allen leerlingen van klas 5 havo en de klassen 5 en 6 atheneum en gymnasium van het Kottenparkcollege te Enschede werden in groepen van 6 verdeeld over de $5 \times 16 = 80$ tekstpassages die Van Remoortele (1974) gebruikte¹⁷. Daarbij werden elke groep $80 : 20 = 4$ testpassages aangeboden. De volgorde waarin we de leerlingen hun testpassages aanboden, was een willekeurige, zodat bevoordeling zoals bedoeld in paragraaf 2.2.1 werd uitgesloten. De leerlingen kregen een instructie, identiek aan de instructie die Van Remoortele (1974:35) gaf. Hieronder vindt u de gemiddelde scores en de standaarddeviaties per tekstfragment met daarnaast de gemiddelde scores en de standaarddeviaties, zoals gevonden door Van Remoortele en weergegeven in Van Hauwermeiren (1975a: 501).

Tekstnummer	Gemiddelde cloze-score (in %)		Standaarddeviatie (in %)	
	door ons gevonden waarden	door Van Hauwermeiren opgegeven waarden	door ons gevonden waarden	door Van Hauwermeiren opgegeven waarden
1	36.1	37.441	10.6	5.07
2	36.9	37.427	13.6	5.31
3	39.4	38.093	7.0	5.77
4	42.4	38.726	11.2	8.29
5	45.3	38.336	12.0	7.00
6	55.2	44.673	13.7	9.62
7	36.1	34.376	12.7	7.59
8	35.6	34.383	14.6	7.82
9	41.7	36.343	14.0	6.79
10	44.6	40.706	16.1	4.49
11	45.6	38.740	12.8	6.39
12	34.4	30.976	12.3	6.47
13	50.9	37.186	11.0	4.16
14	39.8	20.363	10.0	5.76
15	57.0	47.046	15.4	5.97
16	37.2	33.353	11.9	5.21

Tabel 1: Gemiddelde scores en standaarddeviaties uit het oorspronkelijke en het validatie-onderzoek.

Opvallend in tabel 1 is in de eerste plaats dat 'onze' proefpersonen gemiddeld hoger scoorden dan de proefpersonen die Van Remoortele tot zijn beschikking had¹⁸. Daarnaast is opmerkelijk dat de toch al hoge standaarddeviaties die Van Hauwermeiren opgeeft, nog worden overtroffen door de standaarddeviaties die wij hebben gevonden.

Belangrijker echter dan gegevens zoals opgenomen in tabel 1 zijn de gegevens in tabel 2.

Variabele	1	2	3
1			
2	.646		
3	.586	.655	
4	.727	.715	.897

Tabel 2: Correlatiematrix

Variabele 1 = de scores van 'onze' proefpersonen.

Variabele 2 = de scores vermeld in Van Hauwermeiren (1975a).

Variabele 3 = de door F_1 voorspelde scores.

Variabele 4 = de door F_2 voorspelde scores.

Het meest markante gegeven uit tabel 2 is ongetwijfeld de gevonden correlatie tussen de variabelen 1 en 2: deze correlatie (.646) lijkt te laag, wanneer men de omstandigheden beschouwt waarin het onderzoek van Van Remoortele en ons validatie-onderzoek hebben plaatsgevonden: precies dezelfde testversies en gelijksoortige groepen proefpersonen werden gebruikt.

Opvallend is verder dat de coëfficiënten in de tweede kolom van de matrix (.655 en .715), die de correlaties weergeven tussen de cloze-scores waar Van Hauwermeiren zich op baseerde en de voorspellingen van de daarvan afgeleide formules, nogal afwijken van de waarden die Van Hauwermeiren opgeeft voor de multiple R van F_1 en F_2 , resp. .836 en .898. De oorzaak van die verschillen moet wel gelegen zijn in de geringe analyse-betrouwbaarheid van F_1 en F_2 (zie paragraaf 2.2.6); de waarden die wij voor de verschillende linguïstische variabelen in F_1 en F_2 in de criteriumteksten hebben gevonden, moeten (belangrijk) hebben afgeweken van de waarden die Van Hauwermeiren heeft gehanteerd. Zekerheid daarover is echter niet mogelijk, omdat Van Hauwermeiren (1975a), noch Van Remoortele (1974) de in de criteriumteksten gevonden waarden voor die variabelen weergeeft, en het onmogelijk bleek deze waarden anderszins nog te achterhalen.

Vervolgens moeten we nog ingaan op de correlaties tussen variabele 1 enerzijds en de variabelen 3 en 4 anderzijds, de correlaties dus tussen de (door ons) waargenomen scores en de door de formules voorspelde scores. Zowel voor F_1 als F_2 lijken ons deze waarden (resp. .586 en .727) te laag om eruit te concluderen, dat de correlaties die Van Hauwermeiren (1975a) (resp. .836 en .898) vond tussen de testscores en de formulescores geen toevallige correlaties waren.

Rest nog de correlatie tussen de variabelen 3 en 4, de door resp. F_1 en F_2

voorspelde scores. Dat de gevonden waarde (.897) zo hoog is mag geen verbazing wekken. Bezien we immers F_1 en F_2 dan blijkt dat de linguïstische variabelen die in F_1 een rol spelen ook grotendeels de uitkomsten van F_2 bepalen.

Een laatste opvallend correlationeel verband, dat overigens niet in tabel 2 is weergegeven, betreft de door 'onze' proefpersonen behaalde scores en de door de formules van Flesch-Douma en Brouwer (Leesindex A) voorspelde waarden¹⁹. Gezien de methodologische bezwaren die tegen deze formules kunnen worden aangevoerd, konden we geen hoge correlaties verwachten. Het bleek echter dat de voorspellende waarde van deze formules nauwelijks onderdeed voor die van F_1 en F_2 ; de correlatie tussen de voorspellingen van de Flesch-Douma-formule en de scores die 'onze' proefpersonen behaalden, bedroeg .648, in het geval van de Brouwerformule vonden we een correlatie van .606. Ook Zondervan e.a. (1976: 433) delen mee dat ze, tot hun verrassing, in hun validatie-onderzoek hoge correlaties vonden tussen hun criteriumscores en de formules van Flesch-Douma en van Brouwer. Wellicht zijn deze formules, ondanks de gebrekkige constructieprocedure waarmee ze tot stand gekomen zijn, voorlopig nog de minst onbruikbare leesbaarheidsformules voor 'informatieve Nederlandse teksten', bestemd voor een 'algemeen publiek'.

4. Conclusie

Uit de beschrijving van de stand van zaken in het Nederlandse leesbaarheidsonderzoek die we in paragraaf 1 hebben gegeven, bleek dat, behalve F_1 en F_2 , alle tot nu toe ontwikkelde leesbaarheidsformules voor Nederlandse teksten bestemd voor volwassenen, op methodologische gronden van de hand moeten worden gewezen. In paragraaf 2 menen we aangetoond te hebben, dat dit ook geldt voor F_1 en F_2 , een conclusie die wordt ondersteund door de gegevens uit het in paragraaf 3 gerapporteerde validatie-onderzoek.

5. Suggesties voor verder onderzoek

Tot nu toe zijn leesbaarheidsformules voor twee doelen gebruikt: 1 als een soort handleiding voor het schrijven van beter leesbare teksten. De (her-)schrijver manipuleert alleen die linguïstische variabelen in de tekst die in een bepaalde leesbaarheidsformule een rol spelen, in de hoop zo een beter leesbare tekst te verkrijgen en 2 als een instrument om de leesbaarheid van bestaande teksten te voorspellen (of zoals Van Poecke (1976: 12) het uitdrukt: "[om ervoor te] zorgen dat de juiste tekst in de juiste lezershanden komt")²⁰.

In deze paragraaf zullen we de geschiktheid van leesbaarheidsformules als middelen om die doelen te bereiken, vergelijken met andere methoden die voor die doelen zijn en kunnen worden ontwikkeld.

5.1 *Leesbaarheidsformules als handleiding voor het schrijven van beter leesbare teksten*

Gelukkig kan worden geconstateerd dat leesbaarheidsformules, o.a. dank zij de voortdurende waarschuwingen van degenen die ze ontwikkelen, steeds minder vaak voor dit doel worden gebruikt. Het misverstand dat correlationeel verband ook altijd direct causaal verband zou zijn, verliest steeds meer terrein. Steeds

minder schrijvers menen dat ze alleen al bijv. door het terugbrengen van gemiddelde woord- en zinslengte hun tekst beter leesbaar kunnen maken omdat daardoor de uitslag van een leesbaarheidsformule positief wordt beïnvloed.

Een beter middel om valide handleidingen voor het schrijven van leesbare teksten te verkrijgen, lijkt om, zoals Drop (1978) aanbeveelt, op basis van een aantal gangbare noties omtrent stappen in het schrijfsproces een schrijfsprocedure te ontwikkelen en die vervolgens te valideren. Daarbij moet dan worden getoetst of en in hoeverre die procedure uitvoerbaar, effectief en efficiënt is. Met name bij de toetsing van de effectiviteit – de beantwoording van de vraag of en in hoeverre “[de procedure] – ook bij grote aantallen gebruikers – een produkt oplevert dat (binnen hun mogelijkheden) de kwalificatie ‘goed stuk’ verdient” (Drop 1978: 5) – komt de vraag aan de orde welke meetinstrumenten voor tekstkwaliteit of -leesbaarheid het meest geschikt zijn, de vraag dus waar we in 5.2 op in zullen gaan. Een belangrijke verdienste van de methode die Drop voorstelt lijkt ons, dat daarin circulariteit zoveel mogelijk wordt vermeden.

De handleiding voor het schrijven wordt niet langer afgeleid van het instrument waarmee tekstkwaliteit of -leesbaarheid wordt voorspeld.

5.2 Leesbaarheidsformules als instrument om de leesbaarheid van teksten te voorspellen

Ook tegen het gebruik van leesbaarheidsformules als leesbaarheidsvoorspellers hebben we een aantal bedenkingen:

- a. Zolang een bevredigende tekst/leestheorie ontbreekt, lijkt het niet erg vruchtbaar om met toch tamelijk willekeurige toetsen als de cloze procedure of vragentoetsen leesbaarheid te willen meten. Beter lijkt het om leesbaarheid (of liever: tekstkwaliteit) te operationaliseren in termen van oordelen van die lezers voor wie de teksten bedoeld zijn, bijv. via paarsgewijze globale jury-beoordeling. Een andere mogelijkheid lijkt voort bouwen op de onderzoeksresultaten van Langer e.a. (1974). Daaruit blijkt dat mensen die getraind zijn in het beoordelen van teksten op vier dimensies, nl. structuur, stijl, bondigheid en aantrekkelijkheid, na korte tijd in staat zijn betrouwbare en tegen externe criteria gevalideerde oordelen over de kwaliteit van teksten uit te spreken.
- b. Met Van Hauwermeiren (1977: 177) zijn we van mening dat algemene leesbaarheidsformules een geringe gebruikswaarde hebben: “[ze] doen weinig meer dan onze verwachting aangaande de leesbaarheid van teksten bevestigen en kwantificeren”.

Minder algemene formules daarentegen hebben onzes inziens het bezwaar dat de investering die nodig is om ze te ontwikkelen, nauwelijks meer loont. Efficiëntere manieren om iets te weten te komen over de kwaliteitsverschillen tussen teksten die behoren tot een duidelijk omschreven type bedoeld voor een duidelijk omschreven publieksgroep, lijken ons de hierboven genoemde methoden: beoordeling door een steekproef uit de gebruikersgroep en/of beoordeling door getrainde beoordelaars. Verder onderzoek zou zich dan bijv. kunnen richten op de vraag of de resultaten die Langer e.a. (1974) gevonden hebben voor het Duitse taalgebied, ook voor het Nederlandse taalgebied zullen gelden.

Noten

- *. Voor hun hulp bij de totstandkoming van dit artikel danken we Jan van der Staak, Michaël Stechouder, Frits Steenhof, Gea Timmer en Ted Yap. Onze dank gaat ook uit naar leerlingen en leraren van het Kottenparkcollege te Enschede voor hun medewerking bij het onderzoek en naar Paul van Hauwermeiren voor zijn commentaar bij een eerdere versie. Uiteraard zijn wij geheel voor de inhoud van dit artikel verantwoordelijk.
1. In de literatuur worden de termen 'leesbaarheid' en 'begrijpelijkheid' vaak door elkaar gebruikt. Een duidelijke uiteenzetting over de verhouding tussen deze begrippen is te vinden in Van Hauwermeiren (1975b: 5-10). We zullen hier steeds de term 'leesbaarheid', in de zin van 'begrijpelijkheid' gebruiken.
 2. Er komen nl. twee variabelen in voor die .87 correleren; deze formules voldoen dus niet aan de eis van onafhankelijkheid van de verklarende variabelen (zie p. 13).
 3. Een uiteenzetting over de betekenis van de symbolen in F_1 en F_2 vindt u in paragraaf 2.
 4. Ons uitgangspunt daarbij is het leesbaarheidsonderzoek in enge zin, gedefinieerd door Van Poecke (1976: 10) als "[...] deze translinguïstische (of retorische, pragmatische) discipline die het effect van de verwoording van de boodschap op de verstaanbaarheid, en derhalve op het cognitief effect van deze boodschap, bij een bepaald lezerspubliek wil onderzoeken". Deze omschrijving komt overeen met wat Klare (1963: 1) verstaat onder *readability as ease of understanding*. *Readability as legibility* (de typografische leesbaarheid) en *readability as interest* (de mate waarin een tekst belangstelling opwekt) blijven dus buiten beschouwing.
 5. De variabelen 'gemiddelde woordlengte' en 'gemiddelde zinslengte' uit de Flesch-formule bleken in Nederlandse teksten andere waarden aan te nemen dan in vergelijkbare Engelse teksten (zie Brouwer 1976: 492-493).
 6. Een meetmethode die erin bestaat dat men proefpersonen in een tekst weggelaten woorden laat invullen. Meestal wordt daarbij systematisch elk vijfde woord gedeleerd. Hoe meer woorden nu (exact) juist zijn ingevuld, des te beter zou de tekst begrepen zijn. Voor een uitgebreide uiteenzetting zie: o.a. Van Hauwermeiren (1975b: 77-88).
 7. Zoals Van Hauwermeiren (1977:175) opmerkt: "Blijkbaar blijft de 'ver nederlandse' van de R.E.-formule de belangstelling van sommige onderzoekers genieten [...] ondanks het feit dat nu al oorspronkelijke formules bestaan. Is het toeval dat de pedagogische tijdschriften de Douma-Brouwer-Van der Werff-lijn verder doortrekken, hoewel in de taalkundige tijdschriften al een paar oorspronkelijke Nederlandse formules voorgesteld zijn?" Van Hauwermeiren doelt daarbij op de formules van Zondervan, Van Steen en Gunneweg alsmede op F_1 en F_2 die immers wel degelijk op de cloze procedure zijn gebaseerd.
 8. Brouwer namelijk meende een meting van de leesbaarheid van zijn criteriumteksten te kunnen omzeilen door uit te gaan van op intuïties van enkele personen gebaseerde oordelen omtrent de moeilijkheid van de criteriumteksten.
 9. Aspecten genoemd in Van Hauwermeiren (1975b), die o.i. bij de constructie van F_1 en F_2 wel bevredigend zijn behandeld, zullen we hier niet verder bespreken. Het gaat daarbij om de keuze van de linguïstische variabelen die in het onderzoek werden betrokken en de formulevorm.
 10. Zelfs blijkt bij nadere analyse dat de gevonden verschillen zo klein zijn dat slechts in 42% van de gevallen van significante verschillen (op 5% niveau) kan worden gesproken.
 11. Paul van Hauwermeiren was zo vriendelijk om ons in dit verband te wijzen op een doctoraal-scriptie van C. Lutters (Lutters 1976). Daarin worden steekproeven uit vier

oorspronkelijke Nederlandse literaire teksten vergeleken met steekproeven uit vier uit het Engels in het Nederlands vertaalde literaire teksten. Het ging daarbij met name om het aantal woorden in tangconstructies. De conclusie was dat er geen duidelijk verschil bestond tussen de twee categorieën teksten. Gezien echter het beperkte aantal en de (literaire) aard van de onderzochte teksten, lijkt ons generalisering van deze conclusie naar de verzameling *informatieve teksten* niet goed mogelijk.

12. Deze factoren zouden de cloze-score van de laatste test kunnen verlagen; daartegenover staat dat een factor als gewinning of routine de latere tests kan bevoordelen. Men mag er echter niet van uitgaan dat beide groepen factoren elkaar zouden compenseren.

13. Blijkens Van Hauwermeiren (1975a: 511) is bij de constructie van F_1 en F_2 ook zo'n lineair model voorondersteld. Argumenten voor die vooronderstelling (de vooronderstelling dus dat de relaties tussen leesbaarheid en tekstvariabelen rechtlijnig zouden zijn) worden daarbij niet gegeven.

14. In Van Hauwermeiren (1975a: 512) moet in de formule voor de gecorrigeerde multiple R^2 een drukfout gesloten zijn: de formule luidt:

$$R^{-2} = R^2 - \frac{K(1 - R^2)}{T - K - 1} \quad \text{Yamane (1973: 979). Aldaar genoteerd als}$$

$$R^{-2} = -1 - \frac{SE}{SYY} \left(\frac{T-1}{T-K-1} \right) \quad \text{waarbij} \quad \frac{SE}{SYY} = 1 - R^2.$$

Voor de formule voor R^2 , zie bijv. Blalock (1972: 458).

15. Een niet linguïstisch geschoolde gebruiker als Korsten (1975) doet bij de toepassing van de formules dan ook een beroep op een taalkundige.
16. In o.a. Zondervan & Van Caldenborgh (1971-1972: 348) wordt een dergelijk onderzoek aangeduid als de bepaling van de *predictive validity*. De term *cross validity* wordt daar gebruikt om onderzoek mee aan te duiden waarin wordt nagegaan in hoeverre het oorspronkelijk criterium overeenstemt met andere criteria voor de leesbaarheid. Voor de beschrijving van het door ons uitgevoerde onderzoek zullen we hier verder de term *validatie-onderzoek* gebruiken.
17. Oorspronkelijk was daarbij de bedoeling om de indeling in groepen van 6 te baseren op de uitslagen van een voortoets, waarin de 'leesvaardigheid' van de leerlingen zou worden gemeten. De voortoets die we voor dat doel afnamen was een cloze procedure test die bestond uit cloze-versies van 2 teksten (elk ongeveer 250 woorden) van hetzelfde type als de 16 teksten die Van Remoortele (1974) gebruikte. De resultaten van deze voortoets bleken echter onbruikbaar omdat de test-hertest betrouwbaarheid (berekend via de split-half methode, zie bijv. Looijmans e.a. (1978: hst. III en VI) (van het meetinstrument te laag was. De test-hertest betrouwbaarheid bedroeg namelijk slechts .62, opmerkelijk laag vergeleken met de uit de literatuur bekende test-hertest betrouwbaarheidsgegevens van de cloze procedure (zie bijv. Wijnstra (1977) en De Landsheere (1978)).
18. Een mogelijke oorzaak zou kunnen zijn dat de proefpersonen uit het Belgische onderzoek meer moeite hebben met een aantal specifiek Noordnederlandse uitdrukkingen dan onze proefpersonen. Een informeel onderzoek waarin enkele Belgische proefpersonen de test-passages op dit punt vergeleken gaf aanleiding te concluderen dat er geen passages in het onderzoek gebruikt waren die dergelijke uitdrukkingen bevatten.
19. De Flesch-Douma-formule luidt: $RE = 206.84 - 0.77 \, wl - 0.93 \, zl$. De *LI A*-formule van Brouwer luidt: $LI A = 195 - 2/3 \, wl - 2 \, zl$. De symbolen in deze formules moeten als volgt worden geïnterpreteerd: *RE* en *LI A* staan voor de voorspelde leesbaarheid; *wl* staat voor het aantal lettergrepen per 100 woorden, en *zl* staat voor het gemiddeld aantal woorden per zin.

20. Van Poecke (1976) betoogt overigens dat dit niet het enige doel is van het leesbaarheidsonderzoek. Eigenlijk gaat het volgens hem om het streven naar een terugkeer naar een minder conventionele en natuurlijke taal (*Plain Talk*), een streven door Van Poecke aangeduid met de term *impliciet Rousseauisme*.

Bibliografie

- Blalock, H.M., *Social Statistics*, 2nd ed., Tokyo 1972.
- Bormuth, J.R., Experimental applications of cloze tests, in: J. Figurel (Ed.), *Improvement of reading through classroom practice*, Z.p., 1964, p. 303-306.
- Brouwer, R.H.M., Leesbaarheidsformules voor de Nederlandse taal. Een kritische beschouwing, in: *Pedagogisch tijdschrift/Forum voor Opvoedkunde*, 1(1976), p. 491-505.
- Chall, J.S., *Readability: an appraisal of research and application*, Columbus 1958.
- De Landsheere, G., *Le test de closure. Mesure de la lisibilité et de la compréhension*, Paris 1978.
- Drop, W., Validering van een schrijfpprocedure door herschrijving: een begin, in: *De nieuwe taalgids*, 71(1978), p. 1-17.
- Groot, A.D. de, *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*, 3e dr., 's-Gravenhage 1966.
- Hauwermeiren, P. van, Het leesbaarheidsonderzoek met toepassing op Nederlandse teksten, in: *Leuvense Bijdragen* 52 (1963), p. 12-75.
- Hauwermeiren, P. van, Leesbaarheidsformules voor informatieve Nederlandse teksten, in: *Spektator* 4(1974-1975), p. 499-520. (Aangeduid als 'Van Hauwermeiren (1975a)').
- Hauwermeiren, P. van, *Het leesbaarheidsonderzoek*. Groningen 1975 (Aangeduid als 'Van Hauwermeiren (1975b)').
- Hauwermeiren, P. van, De ontwikkeling van Nederlandse leesbaarheidsformules, in: *Tijdschrift voor Massacommunicatie* 5 (1977), p. 161-181.
- Jansen, C.J.M. en E.T. Woudstra, De cloze procedure, een bruikbare maat voor schriftelijke taalvaardigheid?, in: *Tijdschrift voor Onderwijsresearch* 3 (1978), p. 87-89.
- Klare, G.R., *The measurement of readability*, Ames 1963.
- Klooster, W.G., H.J. Verkuyl en J.H.J. Luif, *Inleiding tot de syntaxis. Praktische zinsleer van het Nederlands*, 2e ongew.dr., Culemborg enz. 1970.
- Korsten, A.F.A., Leesbaarheid van Overheidsbrochures en Sociale Ongelijkheid, in: *Tijdschrift voor Massacommunicatie* 3 (1975), p. 90-113.
- Langer, I., F. Schulz von Thun en R. Tausch, *Verständlichkeit*, München 1974.
- Looijmans, P., H. Palm en S. Stijssiger, *Inleiding Onderzoeksmethoden. Test-theorie*, [Utrecht], 1978 (R.U. Utrecht, Instituut De Vooy's).
- Mellenbergh, G.J., *Bekend, maar onbemand (deel I)*, [Amsterdam] 1976, (Universiteit van Amsterdam).
- Poecke, L. van, *Het trouweloze medium. Een onderzoek naar de correctieve retoriek van het leesbaarheidsonderzoek*, Leuven, 1976. (Werkdocumenten van het Centrum voor Communicatiewetenschappen, nr. 6).
- Remoortele, R. van, *Contribution to the measurement of readability of Dutch popular scientific texts translated from English*, Gent, 1974 ((Onuitgegeven) licentiaatsscriptie, Hoger Instituut voor Vertalers en Tolken).
- Rijppma, E. en F. Schuringa, *Nederlandse Spraakkunst*, bewerkt door J. van Bakel, 22e dr., Groningen 1969.
- Wijnstra, J.M., Het gebruik van de cloze procedure als maat voor schriftelijke taalbeheersing, in: *Tijdschrift voor Onderwijsresearch* 2 (1977), p. 262-269.
- Yamane, T., *Statistics. An introductory analysis*. 3rd ed. Tokyo 1973.
- Zondervan, F. en P. van Caldenborgh, Leesbaarheidsformules, constructie en betrouwbaarheid, in: *Spektator* 1 (1971-1972), p. 341-351.
- Zondervan, F., P. van Steen en G. Gunneweg, De leesbaarheid van basisschoolteksten. Objectieve ordeningscriteria voor instructieve teksten, in: *De nieuwe taalgids*, 69(1976), p. 426-445.

De invloed van theoretische ontwikkelingen in de psychologie op de methoden van het onderzoek naar het onthouden van teksten

G. VAN DAM

Samenvatting

Er hebben zich sinds ca. 1960 ingrijpende ontwikkelingen voorgedaan wat betreft de psychologische opvattingen omtrent het leren en het functioneren van het geheugen bij de mens. De grondgedachten van het behaviourisme werden verlaten, om plaats te maken voor opvattingen, waarin het accent ligt op de mens als actief handelend, denkend, oordelend individu. Deze ontwikkeling had tot gevolg, dat men ging zoeken naar nieuwe methoden, om de resultaten van leer- en geheugenactiviteiten te analyseren. Dit wordt geïllustreerd aan de hand van voorbeelden uit het onderzoek naar het leren en onthouden van teksten.

1 Inleiding

Dat mensen in staat zijn om niet onaanzienlijke hoeveelheden informatie in hun geheugen op te slaan en deze later al dan niet op commando weer te reproduceren, is een vaststaand feit. Ook is het duidelijk, dat deze informatie-opslag en reactivering bij mensen gecompliceerder verloopt dan bij elektromechanische apparaten zoals bandrecorders, grammofoons e.d. Een al direct in het oog springend verschil is bijv., dat mensen dingen kunnen vergeten, hetgeen bij mechanisch of elektromagnetisch geregistreerde gegevens niet gebeurt, tenzij speciale ingrepen worden toegepast. Een andere opvallende eigenaardigheid van het functionerende menselijke geheugen is, dat aangeboden gegevens slechts zelden letterlijk worden vastgelegd c.q. gereactiveerd: doorgaans wordt informatie verwerkt, in verband gebracht met andere gegevens, 'vertaald' in andere bewoordingen of zelfs in een andere taal etc.

Dit alles duidt erop, dat het ten aanzien van het menselijk geheugen veel moeilijker zal zijn, precies te voorspellen, hoe het in een bepaalde situatie zal functioneren. Toch betekent dit nu – gelukkig – ook weer niet, dat we wat betreft ons geheugen geheel aan het toeval zijn overgeleverd. Bij verreweg de meeste mensen functioneert hun geheugen dusdanig bevredigend, dat zij zich met behulp ervan goed kunnen handhaven in hun leefsituatie. Hoewel precieze voorspelbaarheid ontbreekt, bestaan er kennelijk toch bepaalde regelmatigheden. Het zijn nu juist dit soort regelmatigheden welke men in de experimentele psychologie tracht te achterhalen.

Een probleem dat zich hierbij onmiddellijk voordoet, is het feit dat de meeste informatie waar mensen in het dagelijks leven zoal mee omgaan, tamelijk complex van karakter is. Met uitzondering van gegevens als losse namen, telefoonnummers e.d. (die vaak het lastigst zijn om te onthouden) bestaat deze informatie doorgaans uit samenstelsels van meerdere gegevens, welke op een bepaalde manier met elkaar in verband staan.

Deze samenstelsels van gegevens worden, zoals we zagen, door mensen bovendien nog 'verwerkt', waarbij persoonlijke karakteristieken van het individu een min of meer belangrijke rol kunnen spelen. Een van de kernproblemen van de